Ikomi, P. A., Boehm-Davis, D. A., Holt, R. W., & Incalcaterra, K. A. (1999). Jump seat observations of advanced crew resource management (ACRM) effectiveness. Proceedings of the Tenth International Symposium on Aviation Psychology. The Ohio State University: Columbus, OH.

JUMP SEAT OBSERVATIONS OF ADVANCED CREW RESOURCE MANAGEMENT (ACRM) EFFECTIVENESS

Philip A. Ikomi, Deborah A. Boehm-Davis, Robert W. Holt, & Kara A. Incalcaterra

George Mason University,

Fairfax, Virginia

ABSTRACT

A proceduralized form of Crew Resource Management (CRM), Advanced CRM (ACRM) was developed and evaluated in a regional airline. Crews in two different fleets were either trained and used ACRM as standard operating procedure (SOP) or were not trained and did not use ACRM as SOP. Jump seat evaluations of ACRM were executed in both fleets on line flights. Fifty line flights were observed by a separate cadre of 5 air pilot observers. Results showed that the performance of the ACRM trained crews was significantly superior to that of the non-ACRM trained crews on 13 of the 20 items evaluated.

BACKGROUND

The jump seat observation study is a subset of a larger study on the proceduralization of crew resource management (CRM) called Advanced Crew Resource Management (ACRM) at a regional airline in the eastern United States. The idea of proceduralizing CRM stems from a consideration that resource management is a critical aspect of job performance. CRM in the aviation domain has focused on issues of crew coordination and communication (Foushee & Helmreich, 1988). Foushee and Helmreich assumed in their work that better crew coordination and communication will yield improved performance. While there has been some evidence that improved communication and coordination does improve crew performance (see Foushee & Manos, 1981), the most effective methods for changing communication and coordination patterns in the cockpit remains unclear.

The approach to changing communication and coordination patterns in this study was to proceduralize key aspects of CRM and institute these procedures as part of fleet SOP. This has been described as fourth generation CRM (Helmreich, Ashleigh, & Wilhelm, 1999). Carrier trainers felt CRM principles were not specified in a sufficiently concrete fashion or connected to specific observable actions that would allow evaluators to accurately assess and possibly fail pilots on CRM aspects of performance when they felt the pilots had not performed adequately. Pilots also felt that the translation of CRM principles to specific cockpit actions was difficult.

Therefore, CRM principles were developed into CRM procedures appropriate for this carrier (see Holt, Boehm-Davis, Hansberger, Beaubien, Incalcaterra, and Seamster, 1998). These procedures were incorporated into fleet SOP and evaluated by specific observable behaviors connected to each procedure. Over a three-year period the research team developed, implemented and evaluated the effects of these procedures. The evaluation in the third year of the project used multiple methods of assessing the effectiveness of the procedures. All evaluation methods compared the ACRM-trained and SOP fleet, in which the crews had received ACRM training in addition to the traditional CRM training, to the control fleet in which the crews had only received traditional CRM training. The evaluation methods were jump seat observations, Line Operational Evaluations, Line Checks, Surveys of Instructor/Evaluators (I/E), and surveys of all pilots in the airline. This report summarizes the evidence from jump seat observations.

METHOD

Subjects

A sample of 50 crews from an eastern United States regional airline served as subjects. The crew as a functional unit was the focal point for the ACRM training and the jump seat evaluations. Crews were recruited when they showed up for flight and had the option of declining participation after the jump seat observer briefed them on the observation strategy.

Ikomi, P. A., Boehm-Davis, D. A., Holt, R. W., & Incalcaterra, K. A. (1999). Jump seat observations of advanced crew resource management (ACRM) effectiveness. Proceedings of the Tenth International Symposium on Aviation Psychology. The Ohio State University: Columbus, OH.

Observation Form

Observation forms were designed to target the occurrence and effects of proceduralized ACRM in normal operations. The form was divided into four sections. The first section contained experience and training items for each crew member while the rest of the form was devoted to the 20 items pertaining to the three phases of flight that were used in the evaluation. The second section contained departure phase items, the third section contained the cruise phase items, and the fourth section contained the arrival phase items.

The Observers

The observers were three pilots from the grant team and two from the regional carrier. All pilots had at least a commercial license with one exception who was a private pilot but with aviation evaluation experience. None of the pilots was a member of the cadre of evaluators for the observed fleets, and all observers emphasized that they were functioning in a research capacity rather than evaluating for the carrier or for the FAA.

Evaluation Procedure

True random assignment of observers to flights was not feasible due to load restrictions, schedule complexities, and limited jump seat availability. Therefore, flights were observed on a next-available basis without any other considerations. The first step was to explain to the crew that observations were non-jeopardy and not connected with FAA or the air carrier requirements. Informed verbal consent for each observation was obtained from the crew prior to take off. Consent was denied on only one occasion.

The three basic phases of flight on which the observation form was based were defined as follows: 1. Departure phase. This phase included all performance from preparations at the gate to arrival at cruise altitude. 2. The cruise phase. This included all performance from arrival at cruise altitude to preparation for descent at the end of cruise. 3. Arrival phase; which included all performance from the preparation for descent to engine shut down. The phases were reminiscent of event sets used in Line Operational Evaluation (LOE) design (Hamman, Seamster, Smith, & Lofaro, 1993). The flights for this regional carrier averaged between 30 minutes and 1.5 hours in duration, so these phases were correspondingly shorter than flights at a major carrier.

The observer carefully watched the crew and the situation during each phase of the flight and completed the observation form as soon as the flight ended. For each phase of flight, components of crew briefs were checked off from a list that included all standard elements and the elements required by ACRM. The quality of crew performance on specific ACRM procedures and related items was also evaluated by detailed items using a 5-point Likert scale response format. Each observer also gave a summary evaluation of crew effectiveness for each phase of flight and for the entire flight.

RESULTS

Preliminary analyses

Crews were rated on their performance during the three phases of flight -- take-off, cruise, and approach/descent – and on their overall flight performance. Preliminary analyses were conducted to compare the rating profiles across the observers. We did this to ensure that none of the observers was skewing the ratings with overly positive or overly negative ratings compared to the group ratings. We compared the percentage of observed versus non-observed ratings given by each observer for each dichotomous item. For items scored on a scale, we computed the relative percentage of 1's, 2's, 3's, 4's and 5's by each observer as well as the percentage of items left blank by each observer. We found no clear discrepancies among the observers' rating profiles.

Detailed analyses

The ACRM fleet crews were compared with the non ACRM fleet crews on their flight performance using a two-sample design . We used independent sample *t*-test to compare differences between ACRM and the non-ACRM fleets on each of the twenty items scored on a scale. These individual item analyses were designed to detect fine-grained differences between the two experimental fleets if any existed. The *t*-tests revealed that in 13 of the 20 items, crews from the ACRM trained and SOP fleet showed superior performance compared to

Ikomi, P. A., Boehm-Davis, D. A., Holt, R. W., & Incalcaterra, K. A. (1999). Jump seat observations of advanced crew resource management (ACRM) effectiveness. Proceedings of the Tenth International Symposium on Aviation Psychology. The Ohio State University: Columbus, OH.

crews from the control fleet. These results are shown in Table 1. Fleet differences in performance shown with the t-test results are further highlighted by phase of flight in Figs. 1-3.

Positive Differences. Five of the items showing positive differences between the fleets were focused on the thoroughness or timing of briefings (items 1, 4, 9, 13, and 15). This result is plausible since the ACRM procedures for normal operating conditions stressed customizing briefing content to the flight situation and executing briefings during low-workload periods of the flight. Two items (8 and 14) showing positive differences focused on the crew avoiding distractions in the cruise and arrival phases of flight. On the arrival phase, two additional items showing a positive difference were being mentally prepared for critical events or issues in the arrival phase (17) and formulating and communicating bottom lines and backup plans for the arrival (18). The differences on these items is plausible because establishing bottom lines and backup plans is a required part of ACRM procedure, and the requirement to do this look-ahead and planning may have been reflected in better overall preparation for unanticipated events on the arrival. The final four items showing a positive difference concerned the overall effectiveness of the crew during departure, cruise, and approach/descent phases of flight, and the overall effectiveness item referring to all phases of flight.

No Differences. Three of the items showing no significant positive difference concerned the crew being organized and prepared during the departure, cruise, and arrival phases of flight (items 5, 10, and 16, respectively). Although interpreting this lack of effect is extremely speculative, the consistency of finding no effect on this item content may reflect the reality that the ACRM procedures did not strongly affect the organization and preparation of crews on the line. The other four items showing no differences concerned mental preparation of the crew for the departure phase (Item 2), avoiding distractions during the departure (Item 3), communicating the status of the aircraft during cruise (Item 7), and effectively dividing workload during arrival (Item 12). It should be noted that some of the differences on these items were in fact positive but not statistically significant with the limited sample size.

Table 1

Fleet differences in performance for specific observation items as rated by jumpseat observers.

Item:	Mean Untrnd Fleet	Mean ACRM Fleet	t-value
Departure Phase items:			
1. Crew gave tailored CLEARANCE brief according to the particular flight conditions.	n=15 2.93	n=17 4.12	3.25**
2. Crew mentally prepared for critical issues for departure phase of flight.	n=19 3.78	n=26 4.15	1.68
3. Crew avoided distraction during departure phase of flight.	n=19 4.05	n=26 4.08	0.11
4. Normal briefs were conducted during low workload periods of flight.	n=19 2.95	n=26 4.23	3.81**
5. Crew was organized and prepared in the cockpit	n=19 4.05	n=26 4.31	0.98
6. Overall crew performance for the departure phase of flight.	n=20 2.92	n=27 3.52	2.90**

Cruise Phase items:			
7. The captain and first officer communicated the status of the aircraft and relevant conditions	n=18 3.55	n=23 4.08	1.7
8. Crew avoided distractions during the cruise phase	n=19 3.21	n=27 3.78	1.98*
9. Normal briefs were conducted during low workload periods of flight.	n=18 3.83	n=25 4.36	2.14*
10. Crew was organized and prepared in the cockpit	n=19 4.05	n=27 4.15	0.42
11. Overall crew performance during cruise phase of flight.	n=20 2.95	n=27 3.46	2.52*
Arrival Phase items:			
12. Workload was effectively divided to manage situation.	n=17 3.29	n=25 3.72	0.97
13. Crew briefed or discussed the particular flight situation for arrival.	n=20 3.55	n=27 4.33	3.30**
14. Crews avoided distractions during the arrival phase of flight.	n=20 3.35	n=26 4.35	4.01**
15. Normal briefs were conducted during low work-load periods of flight.	n=18 3.00	n=25 4.08	3.19**
16. Crew was organized and prepared in the cockpit	n=19 4.05	n=26 4.31	1.29
17. Crew was mentally prepared for critical issues for arrival phase of flight.	n=20 3.50	n=27 4.26	2.73**
18. Crew was formulating and communicating bottom lines and back-up plans as needed during arrival phase	se n=20 2.35	n=25 4.08	4.64**
19. Overall crew performance for arrival phase of flight.	n=20 2.73	n=24 3.69	4.71**
Overall Performance Item:			
20. Overall performance for all phases of flight	n=18 2.75	n=25 3.48	3.60**

** $p < .01$
* $p < .05$

Effect size. As shown in Table 1, the observed differences are generally significant at the .01 level. Furthermore, the sizes of the mean differences across these items are fairly large. Across all items, the difference in the average score is .70, which is quite large considering that the items were scored on 5-point scales. Strength of effect as measured by a squared point-biserial correlation indicated that the average amount of adjusted variance explained by the fleet differences on the 13 items was 22.9 %. For the significant items, the jump seat observers noticed fairly large differences that cannot be ascribed to chance variations in the ratings.

General Analyses

A principal components analysis was conducted to see if the items measured on the jump seat observations constituted one or more coherent factors of crew performance. Mean substitution was used for missing data. The results of this analysis showed three components or factors. Factor scores were obtained for each of the preceding factors. Each factor score was analyzed with the same two-sample t-test as used for the item-level analyses. The results of these tests showed that on two of the factors, the fleets were significantly different, $p < .004$ while on the third factor, the fleets were not significantly different.

Finally, a combined, unit-weighted average of all scale items on this questionnaire was formed to represent a global performance composite. A t-test of this global performance index indicated that the trained fleet was significantly better overall, $t = 4.23$, $p < .05$.

DISCUSSION

That the majority of scale-based items were significant suggests that ACRM training plus its incorporation as SOP had an impact on this sample of line flight operations. This conclusion is supported by the general analysis results of the performance factors based on those items. This conclusion is also supported by the congruent evidence of fleet differences reported from the other ACRM evaluation methods (see Holt, et al., 1998).

However, since the trained fleet had both ACRM and ACRM as SOP, these differences in performance could be attributed to the ACRM training, the ACRM SOP implementation, or the combination of these two. It is possible that some of these performance differences were due to the training, others to the SOP implementation, and still others to the combination of the two. Further study is therefore recommended to investigate the reason for the difference in performance between the fleet that received ACRM training and used ACRM as SOP versus the fleet that had only traditional CRM training and did not use ACRM as SOP. One experimental design for such a study would be to measure the performance of a control group of pilots with traditional CRM only (no ACRM), in a fleet where ACRM is not SOP. One experimental group would be made up of ACRM-trained pilots operating in a fleet where ACRM is not SOP. A second experimental group of pilots would have no ACRM training but would operate in a fleet where ACRM is SOP. The difference in performance between the control group and the first experimental group would indicate the effect of ACRM training. Conversely, the difference between the performance of the control group and the second experimental group would be indicative of the effect of ACRM as SOP. We expect that there will be a cumulative positive effect of both manipulations. That is, each experimental group would be better than the control group, although they may be better on different specific items. If this were the case, one could recommend the use of ACRM training and ACRM as SOP because of the advantages of each ACRM intervention.

Another area of study is to investigate in more detail what the items of the jump seat observation form are actually measuring. For instance, we are already looking into the factors that the items represent in terms of the principles of CRM. For instance, what items measure situational awareness, and which ones address communication or leadership? Knowing what these items are would enhance our understanding of the specific effects of ACRM integration into the SOP of an airline and make for better confidence on the part of pilots on the adoption of ACRM procedures in their operations. This information may, for example, clarify why there was a consistent pattern of positive results for most items, but a consistent lack of effect for the "organization and preparation" items

REFERENCES

Foushee, H. C. & Helmreich, R. L. (1988). Group interaction and flight crew performance. In E. L. Wiener & D. C., Nagel (Eds.), Human factors in aviation (pp. 189-227). San Diego, CA: Academic Press.

Foushee, H. C. & Manos, K. L. (1981). Information transfer within the cockpit: Problems in intracockpit communications. In C. E. Billings & E. S. Cheaney (Eds.), Information transfer problems in the aviation system (NASA Technical Paper 1875; pp. 63-71). Moffett Field, CA: NASA-Ames Research Center.

Hamman, W. R., Seamster, T. L., Smith, K. M., & Lofaro, R. J. (1993). The future of LOFT scenario design and validation. Proceedings of the Seventh International Symposium on Aviation Psychology. Columbus, OH: The Ohio State University.

Helmreich, R. L., Merritt, A. C., & Wilhelm, J. A. (1999). The evolution of crew resource management training in commercial aviation. The International Journal of Aviation Psychology, 9(1), 19-32.

Holt, R. W., Boehm-Davis, D. A. Hansberger, J. T., Beaubien, J. M., Incalcaterra, K. A., & Seamster, T. L. (1998). Evaluation of proceduralized CRM training in a regional airline – Final Report. , Fairfax VA: FAA Research Team, George Mason University ARCH Lab.

Ikomi, P. A., Boehm-Davis, D. A., Holt, R. W., & Incalcaterra, K. A. (1999). Jump seat observations of advanced crew resource management (ACRM) effectiveness. Proceedings of the Tenth International Symposium on Aviation Psychology. The Ohio State University: Columbus, OH.

Figure 1: ACRM effects on departure.

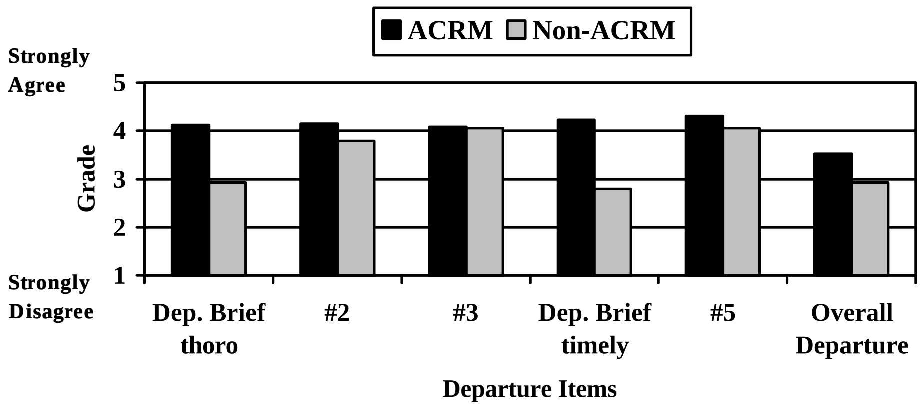

Figure 2: ACRM effects on cruise

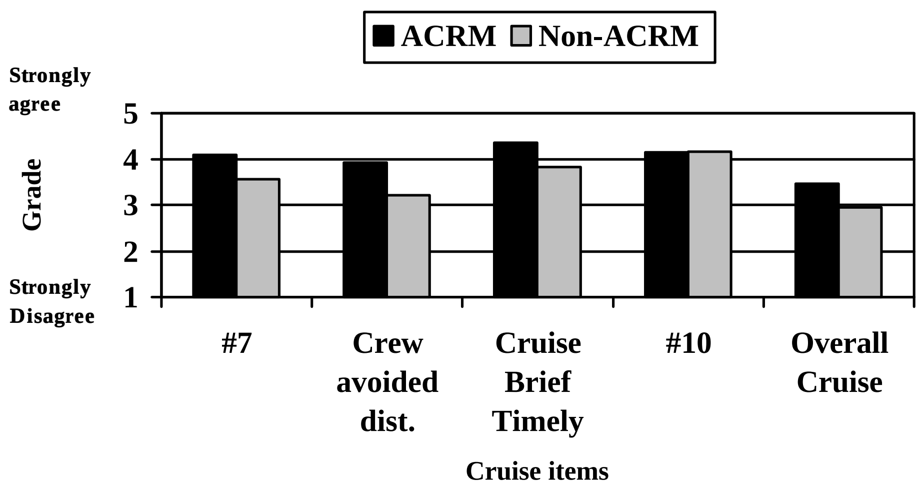

Figure 3: ACRM effects on arrival